# Mathematization in Synthetic Biology: Analogies, Templates, and Fictions

**Chapter** · April 2017
DOI: 10.1007/978-3-319-54469-4_3

**2 authors:**

Tarja Knuuttila
University of Vienna
**94** PUBLICATIONS   **1,361** CITATIONS

SEE PROFILE

Andrea Loettgers
University of Vienna
**34** PUBLICATIONS   **283** CITATIONS

SEE PROFILE

**Some of the authors of this publication are also working on these related projects:**

Project  Ontologisch Status of Organizationsl Principles in Biology  View project

Project  Possible Life: The Philosophical Significance of Extending Biology (ERC)  View project

# Mathematization in Synthetic Biology: Analogies, Templates, and Fictions

**Tarja Knuuttila and Andrea Loettgers**

*T. Knuuttila*
*University of South Carolina, University of Helsinki*
901 Sumter St., Byrnes Suite, Columbia, S.C., 29208 USA
Email: knuuttil@mailbox.sc.edu, tarja.knuuttila@helsinki.fi

*A. Loettgers*
*University of Geneva, University of Bern*
Rue de Candolle 2, CH-1211 Genève 4, Switzerland
Email: andrea.loettgers@unige.ch, andrea.loettgers@csh.unibe.ch

## 1. Introduction

In his famous article "The Unreasonable Effectiveness of Mathematics in the Natural Sciences" Eugen Wigner argues for a unique tie between mathematics and physics, invoking even religious language: "The miracle of the appropriateness of the language of mathematics for the formulation of the laws of physics is a wonderful gift which we neither understand nor deserve" (Wigner 1960: 1). The possible existence of such a unique match between mathematics and physics has been extensively discussed by philosophers and historians of mathematics (Bangu 2012; Colyvan 2001; Humphreys 2004; Pincock 2012; Putman 1975; Steiner 1998). Whatever the merits of this claim are, a further question can be posed with regard to mathematization in science more generally: What happens when we leave the area of theories and laws of physics and move over to the realm of mathematical modeling in interdisciplinary contexts? Namely, in modeling the phenomena specific to biology or economics, for instance, scientists often use methods that have their origin in physics. How is this kind of mathematical modeling justified?

In the following we will shed light on these questions by focusing on the interdisciplinary research practice of synthetic biology. Synthetic biology is a relatively novel field of research located at the interface of physics, biology, engineering, and

computer science. Being situated in this rather complex disciplinary environment makes model building in synthetic biology a highly interdisciplinary task: Methods, techniques, strategies, and concepts from various, even distant fields enter into and get intertwined in the modeling practice of synthetic biology. One unique characteristic of this practice is due to how synthetic biologists combine various kinds of models: model organisms, mathematical models and synthetic models. The latter ones comprise a novel type of model that are constructed from biological components such as genes and proteins on the basis of mathematical modeling. To understand the rationale of this combinational modeling approach one needs to take a closer look at the strategies of mathematization in synthetic biology.

We will discuss two interrelated means through which synthetic biologists study models of gene regulatory networks: analogies and mathematical templates. Synthetic biologists, we argue, proceed to mathematize gene regulatory networks by compound analogies that draw inspiration from engineered artifacts on the one hand, and model systems with non-linear dynamics on the other hand. Engineered artifacts provide material analogs for biological systems, whereas the theory of complex systems offers formal analogs in the form of various mathematical templates for analyzing oscillatory phenomena.[1] A kind of patchwork model results from such compound analogies that, as we will discuss, consists of elements that may even draw into opposite directions. These mathematical models nevertheless allow synthetic biologists to conceptualize biological regulation in terms of positive and negative feedback loops side-by-side with mathematical templates and methods that have been applied in various contexts dealing with rhythmic/cyclic behavior resulting from non-linear dynamics (e.g., physics, chemical kinetics, ecology, economics). In the modeling process the general templates for describing various forms of interaction are adjusted to the subject matter in question, but they remain nevertheless rather abstract, lacking many known empirical details. This contributes to one typical problem constraining the use of template-based mathematical modeling: such models are usually underdetermined by data.

Interestingly, this does not worry synthetic biologists too much. One reason may be that they do not consider their models to be representations of any specific naturally

---

[1] For a discussion on material and formal analogies, see Hesse (1966), and Knuuttila and Loettgers (2014).

occurring gene regulatory networks. Instead, they consider themselves to be in the business of studying general *design principles* or *network motifs* of gene regulatory systems (i.e., genetic circuits). Being very aware of the fact that the conceptual and mathematical means they use are often transferred from other disciplines, they consider their models to depict only possible mechanisms underlying biological regulation. Such principles could have evolved in natural systems but biological systems might have implemented different kinds of regulatory mechanisms. As a consequence the design principles studied mathematically are best conceived of as fictions and their very fictionality has led synthetic biologists to construct *synthetic models* on the basis of mathematical models. Built from biological material, synthetic models can be considered as experimental objects constructed to study the assumptions and credibility of mathematical models. Yet, as we will show, the relationship between mathematical modeling and synthetic modeling is anything but direct—and synthetic models themselves can also be regarded as fictions, albeit concrete ones. In what follows we will first briefly introduce the field of synthetic biology and then go over to the discussion of the means and process of mathematization in this particular field.

## 2. Synthetic Biology: A Nascent Interdisciplinary Field

Synthetic biology focuses on the design and construction of novel biological functions and systems. It is often understood in terms of the pursuit for well–characterized biological parts to create synthetic wholes, and as such it has typically been understood as a kind of engineering science in which engineering principles are applied to biology (Church 2005). This view is shared by the public understanding of synthetic biology as well as the practitioners themselves. According to Jim Collins, who introduced one of the first synthetic networks, a toggle-switch, in 2000: "[…] synthetic biology was born with the broad goal of engineering or 'wiring' biological circuitry – be it genetic, protein, viral, pathway or genomic – for manifesting logical forms of cellular control" (Khalil and Collins 2010).

However, a more basic science oriented branch of synthetic biology has developed alongside the more engineering and application oriented approaches. This

basic science oriented branch of synthetic biology targets our understanding of biological organization by probing the basic design principles of life by various strategies of modeling (see above). The design and exploration of synthetic models, i.e. engineered genetic circuits constructed from biological material and implemented in natural cell environment, provides the most recent strategy of this kind of approach (Sprinzak and Elowitz 2005).

The two branches of synthetic biology are not isolated but overlap and interact in several important ways. First, both make use of compound analogies to engineered artifacts and abstract model systems showing rhythmic/cyclic behavior. Second, the scientists in both branches employ largely the same theoretical tools and techniques. Third, the results gained in the basic science approach are utilized by the engineering oriented branch and the other way around. The main differences between the two branches thus lie in the primary aims of the scientists working in them, that is, whether they probe design principles in order to learn about the mechanisms operating in biological organisms or search of design principles, which could be used in the engineering of *novel* biological parts and systems. These differences in aims are largely rooted in the different scientific backgrounds of the scientists. For example, the majority of scientists belonging to the first group probing the basic design principles of biological organization comes from physics, whereas most of the scientists belonging to the second, more application oriented group has a background in engineering. Moreover, the synthetic genetic circuits are so far largely in their proof of principle phase, and the actual applications of synthetic biology, like the synthetic malaria drug artemisinin, have in contrast resulted from laborious tinkering processes in the lab.

*Modeling Biological Mechanisms*

Biological systems have an inherent complexity given by the number of the different components and their interactions embodied by them. Metabolic and gene regulatory networks provide examples of biological systems that are extensively studied in synthetic biology (e.g., Bujara, Schümperli, Pellaux, Heinemann, and Panke 2011; Zhang, Rodriquez, and Keasling 2011; Elowitz and Leibler 2000; Nandagopal and Elowitz

2011). In what follows, we will focus on gene regulatory networks. Such networks consist of interacting genes and proteins. Genes and proteins interact via transcription and translation processes. Figure (1) shows a simplified picture of the main steps of these translation and transcription processes. Following the central dogma of molecular biology the DNA (deoxyribonucleic acid) carries all the genetic instructions necessary for the development, reproduction and functioning of an organism. The information stored in DNA is transcribed in the process of RNA (ribonucleic acid) synthesis into individual transportable cassettes, the so-called messenger RNA (mRNA). The individual cassettes carrying the blueprint of a protein as sequences of amino-acids, leave the nucleus and enter a complex protein machinery, the ribosome. In this machinery the transcribed information is translated and used in the formation of the protein.
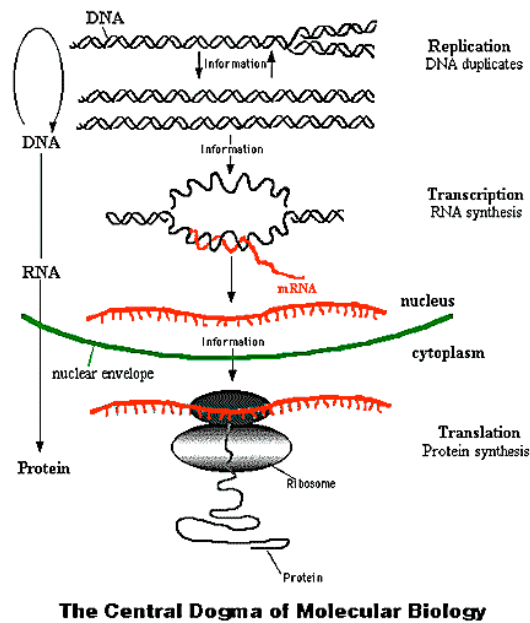


Figure 1. The diagram shows the main elements of the transcription and translation processes according to what is called 'the central dogma of molecular biology'.[2]

The transcription process is activated or inhibited by so-called transcription factors. These are proteins binding to the promoter site of the gene. Figure (2) shows an example of such an activator binding to the promoter site of a gene. In the upper part of the picture the transcription factors binding to the promoter site inhibit RNA polymerase. In the lower part of the picture the transcription factor gets released by proteins moving into the cell and the transcription process starts.
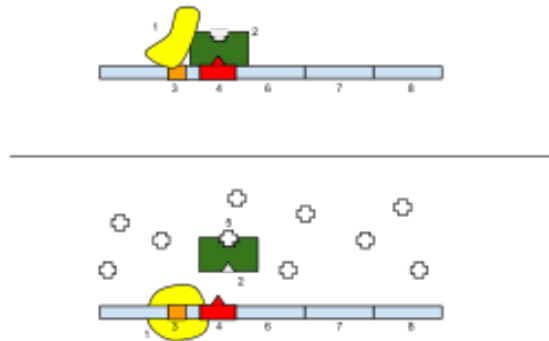


Figure 2. A transcription factor binding at the promoter site and inhibiting RNA polymerase.[3] Only when the transcription factor is released from the promoter site does the transcription process start.

Many important biological functions are based on gene regulatory networks. A prominent example is the circadian clock, which regulates day and night rhythm in biological organisms. The early modelers of biological organization had suggested already in the 1960s that the rhythmic behavior observed in the circadian clock is controlled by a molecular feedback mechanism (Goodwin 1963; Winfree 1967). For example, Colin Pittendrigh, who studied circadian rhythms on *Drosophila* wrote that the: "[...] commonest device in installing regulators—from the control of heartbeat to that of protein synthesis—is negative feedback. And one of the innate tendencies of such

---

[3] (https://en.wikipedia.org/wiki/Promoter_(genetics)) The numbers represent the following entities:1: RNA Polymerase, 2: Repressor, 3: Promoter, 4: Operator, 5: Lactose, 6: lacZ, 7: lacY, 8: lacA.

feedback systems is to oscillate" (Pittendrigh 1961: 122). Yet it has remained an open question as to whether gene regulatory networks in biological systems implement control in the same way as human engineered systems.

Synthetic biologists have followed the tradition of modeling the organization of biological systems in terms of feedback systems, although the relationship of this modeling paradigm to experimental results is far from straightforward. The exploration of regulatory networks in model organisms is very complicated even in the case of such "simple" organisms as the bacteria *Escherichia coli*. It requires a lot of experience and skill to determine the constituent elements (genes, proteins) of the network, its structure and the interaction between the elements. Although the results of experimentation with model organisms are interpreted in terms of design principles adapted from engineering, recent results in synthetic biology show that gene regulatory networks can function in rather counter-intuitive ways. Nature seems to make use of different kinds of principles than human engineers, exploiting, for example, stochastic fluctuations (i.e., noise) in a functional way (cf. Çağatay, Turcotte, Elowitz, Garcia-Ojalvo, and Suel 2009). Engineers typically try to eliminate noise from their systems (see below). Such results as these are bound to question the basic concepts and assumptions made by mathematical modeling of genetic circuits. This friction between the work on mathematical models vis-à-vis model organisms has led synthetic biologists to introduce a novel, additional model type, a synthetic model, which is located between mathematical models and model organisms. In the next sections we will study how the mathematical models of gene regulatory systems are constructed and the way these mathematical models are related to synthetic models.

**3. Analogical Reasoning and the Use of Templates**

As discussed above, scientists have assumed for some time that negative and/or positive feedback mechanisms play an important role in controlling biological functions (Jacob and Monod 1961; Goodwin 1963; Winfree 2001). This assumption was to a large part based on drawing analogies to engineered systems and it also formed the basic idea on which the mathematical models of biological regulation were built. The starting point of such mathematical model consists often of what synthetic biologists call a *toy model*. It is

a model of a stylized abstract mechanism, such as a simple negative feedback mechanism, which is then being extended and refined taking into consideration some subject specific empirical knowledge. Below we will discuss some essential steps of designing a mathematical model from an initial toy model.

Negative feedback loops provide a control mechanism of a very general character: models of negative feedback can be found from many different contexts such as engineering, biochemistry, and physics. In a negative feedback loop the output feeds back into the input of the system repressing the further output and by doing so stabilizes the system. In designing mathematical models of gene regulatory networks an oft-used motif is autorepression. In the case of autorepression, the gene product of a gene A suppresses its own function by binding to its (transcription) site.

The process of autorepression in gene regulatory networks is shown on the left hand side of the picture and the right hand side shows a sketch of a positive feedback loop.



Figure 3. Two network motifs: **b** depicts a negative feedback loop and **c** a positive feedback loop (Alon 2006: 451).

The simple diagram (Figure 3) omits all biochemical details, the structure of genes and proteins as well as such essential parts of the mechanism as the binding of the activator to the promoter site etc. As discussed in the last section, the gene regulatory mechanism is comprised of a transcription and a translation part. During transcription the protein functions as a transcription factor binding to the transcription site of the gene. In the first step one observes an initial rise in the production of the gene product. But when the concentration reaches the repression threshold, which means that the transcription of the

8

gene product becomes repressed, the production rate decreases and the system locks into a steady-state level. This locking into a steady state can be accompanied with oscillations in the protein level. Finding the conditions for sustained oscillations is one of the aims of mathematical modeling because many biological phenomena are periodic/rhythmic, and oscillations are thought to underlie the organization of such important gene regulatory systems as the circadian clock.

However, such oscillations also mark an important difference between feedback mechanisms in engineering and biology. Whereas oscillations in protein levels are essential for controlling biological rhythms, in engineered artifacts oscillations are typically regarded as unwanted and the systems are designed in such a way that oscillations are suppressed. For example, all kinds of electronic control systems typically have to avoid such oscillations in order to function reliably. Familiar examples of such devices are thermostats and cruise controls. This shows that the dynamic features of the regulation mechanisms are different in the case of biological systems despite the initial analogy to engineering. Brian Goodwin described this point in his influential book *Temporal Organization in Cells* (1963) in the following way: "The appearance of such oscillations is very common in feedback control systems. Engineers call them parasitic oscillations because they use up a lot of energy. They are usually regarded as undesirable and the control system is nearly always designed, if possible, to eliminate them" (Goodwin 1963: 5).

Once the simple sketch of a feedback mechanism has been designed, it has to be translated into a mathematical model. Such mathematical model typically consists of a set of differential equations, one modeling the production and degradation of a protein and a second one modeling the mRNA synthesis and degradation. The model then gets adjusted to the particularities of the biological system under study. This modeling approach is very common in kinetic theory and the differential equations are essentially kinetic equations. The challenge is how to choose the relevant biochemical parameters and to determine their values. These limitations are of both of the practical nature (i.e., how to measure the values of the biochemical parameters, which are part of dynamical processes), as well as theoretical, regarding the lack of knowledge and theoretical insight that would guide the search for the most relevant parameters.

Clearly, this process of model construction disregards most biochemical details as well as the rich structures of genes and proteins. Goodwin discussed this abstract character of mathematization accordingly: "[...] in the study of the dynamic properties of a class of biological phenomena such as we are attempting, it is necessary to extract a manageable number of variables from the very large array which occurs in biological system. Is it ever possible to make such an extraction or simplification without doing violence to the very basis of biological organization, its inherent complexity? There is certainly no *a priori* answer to this question, and the only procedure is to try to find some set of variables which appear to constitute a reasonably self-contained system[4] and see if one can get meaningful and useful results relating to its behavior." (Goodwin 1963: 9). The work by Goodwin on temporal organization in cells has been fundamental in modeling cyclic processes in biological systems such as the circadian clock. He provided the elementary mathematical model that functioned as a basic template for the construction of such synthetic models as the *Repressilator*, nearly four decades later.

Let us finally note how the quote by Goodwin mediates the lingering sense of not knowing much of the details. And even after the 1980s when experimental data on genes and proteins involved in circadian clocks in various model organisms started to accumulate, the situation has not changed too much. The limits to what the scientists know about the components, organization and biochemical details of biological systems such as the circadian clock are still pressing. Because of this the already established conceptual frameworks from other areas, such as negative and positive feedback loops, provide at least a starting point for the first modeling attempts. And a corresponding mathematical framework is provided by the computational templates and methods that are used in modeling non-linear dynamic systems. These initial mathematical models for representing and studying various kinds of abstract feedback systems need mathematical articulation and adjustment in view of the systems at hand, yet modelers need to simultaneously take into account the mathematical constraints on how much detailed information can be expressed and studied by these models.

---

[4] This notion of a "reasonably self-contained system" bears an interesting link to the theme of fiction discussed below in section 5.

*Mathematical Templates for the Study of Gene Regulation*

The general equations used by many systems and synthetic biologists (e.g., Goodwin 1963; Elowitz and Leibler 2000; Alon 2006) to describe the processes of transcription and translation are of the following form:

$$\frac{dm}{dt} = \alpha_m - \beta_m \cdot m$$

$$\tag{1}$$

$$\frac{dp}{dt} = \alpha_p - \beta_p \cdot p$$

where m is the concentration of RNA and p the concentration of protein , $\alpha_m$ and $\alpha_p$ the production, and $\beta_m$ and $\beta_p$ the degradation rates of RNA and the protein. This set of differential equations is called rate equations. It is used in chemical kinetics to describe the rate of reactions. These equations provide an example of what Paul Humphreys (2004) calls a computational template. With a concept of a computational template Humphreys refers to genuinely cross-disciplinary computational devices, such as functions, sets of equations, and computational methods, which can be applied to different problems in various domains. An example of such a template is the Lotka-Volterra model, which provides one of the simplest templates for modeling non-linear dynamics. In fact, the rate equations are at a formal level close to Lotka-Volterra equations.[5] The equations are of such a general character that without knowing that they are describing transcription and translation in a genetic network one could as well take them to describe something else, for instance some chemical reaction. In other words, these differential equations are able to describe the general dynamic behavior of various kinds of systems independently from many particularities of these systems. In addition to generality, Humphreys explains the cross-disciplinary usability of computational templates by their tractability. This is one important reason for the introduction of the rate

---

[5] For Lotka-Volterra equations as computational templates, see Knuuttila and Loettgers (2011, 2016).

equations from chemical kinetics to the study of genetic regulation: one can easily calculate the steady states of the system.

The steady states are calculated in the following way:

$$\frac{dm}{dt} = 0$$

$$\frac{dp}{dt} = 0$$

the condition for the steady state is fulfilled by:

$$0 = \alpha_m - \beta_m \cdot m$$
$$0 = \alpha_p - \beta_p \cdot p$$

with:

$$m = \frac{\alpha_m}{\beta_m}$$

$$p = \frac{\alpha_p}{\beta_p}$$

On the basis of these general equations it is possible to specify in more detail the kind of regulation process one is going to study. For example, if the protein p in the set of general equations (1) functions as a repressor one has the case of negative autoregulation/ negative feedback loop. In this case the protein p inhibits the transcription process and therefore its own production. This will lead to oscillations in the protein level.

A first possible step in the adjustment of the differential equations consists in making the assumption that RNAp (RNA polymerase) binds fast to the transcription site being represented by the promoter activity function. This simplifies the problem in such a way that one does not need to take explicitly into consideration the binding of RNAp.[6]

---

[6] Other scientists such as Brian Goodwin take the binding of the RNAp into account. This makes the differential equations more difficult by adding a further variable.

The differential equations for the process of autorepression are then of the following form:

$$\frac{dm}{dt} = \alpha_m \cdot g_R(r) - \beta_m \cdot m$$

$$\frac{dp}{dt} = \alpha_p \cdot m - \beta_p \cdot r$$

with $g_R(r)$ as the promoter activity function and r the number of repressors. The differential equations are non-linear and coupled. The change in the number of m (mRNA) depends on the number of the repressors r, and the other way around, the number of repressors on the number of mRNA. The resulting set of non-linear coupled differential equations cannot be solved analytically.

In sum, in mathematizing biological circuits systems synthetic biologists typically start from the analogies drawn to electric circuits and render the network motifs that describe various kinds of feedback loops into equations by using the toolbox of modeling complex systems, especially the non-linear differential equations. However, as a result the models arrived at are abstract in that they lack a lot of details, and furthermore, there is the problem that the formalisms have not typically been developed with biological systems in mind, although they have been adjusted to take into account some features of the subject matter in question. This abstract, hypothetical and interdisciplinary nature of the mathematical models of genetic circuits has led synthetic biologists to develop a novel modeling method, synthetic modeling. Synthetic models probe to what extent it is legitimate to suppose that gene regulatory networks function on the basis of feedback mechanisms of some kind. Synthetic models are biological networks that are engineered from genes and proteins on the basis of mathematical models. In that sense they can be considered epistemic tools that are constructed to study the design principles depicted by the mathematical models.[7] In a sense this strategy can be seen as a way to materially embody and recontextualize the template-based, sparse and "foreign" mathematical

---

[7] On the notion of an epistemic tool, see Knuuttila (2011).

models into the actual biological conditions where the dynamic, mechanism or function under study is located.

## 4. Synthetic Modeling - The *Repressilator*

The *Repressilator* is a simple engineered gene regulatory network. It is one of the first and most famous synthetic models, introduced in 2000 by Michael Elowitz and Stanislas Leibler (Elowitz and Leibler 2000). The *Repressilator* consists of three interacting genes connected via a negative feedback loop creating oscillations in the protein level. In gene regulatory systems, as we have seen, oscillating proteins are the essential part of the control. The basic network design is taken from electronics: *The Repressilator* is a biological version of a ring oscillator. Before the *Repressilator* was built, Elowitz and Leibler designed a mathematical model of it utilizing mathematical tools that had been developed to study the biological feedback systems (discussed in the previous section).[8] One particular book was of special importance for the design of the *Repressilator*: *Biological Feedback* by Thomas and D'Ari (1990), which presents a formal methodology for analyzing the dynamic behavior of complex systems.[9] In this book feedback systems are analyzed and described in a very general way—that is, it provides computational templates for analyzing different kinds of feedback systems.

*The Mathematical Model of the* Repressilator

Already the seemingly simple set of differential equations presented in the section 3 leads to complex dynamics. More complicated models of gene regulation can be built on this basic template. The mathematical model underlying the *Repressilator* provides an example of such a model (Elowitz and Leibler 2000). The *Repressilator* consists in three genes, *TetR*, *Lacl* and $\lambda cl$, which are arranged in such a way that they inhibit each others'

---

[8] For example, the properties and dynamic features of network motifs describing recurrent structures in genetic networks (e.g. feed-forward and feed-back loops) can be analyzed by making use of the Michaelis-Menten equations (Berg, Tymoczko and Stryer 2002).
[9] Personal communication by Michael Elowitz.

activity (see Figure 4). The fourth gene used in the construction of the *Repressilator* is a Green Fluorescent Protein (GFP). The GFP gene is not part of the differential equations, as it does not contribute to the dynamic of the system (as discussed below). The dynamic of the *Repressilator* results from the following mechanism: The protein related to each gene represses the protein production of its neighboring gene. This leads to oscillations in the protein levels of the respective genes.
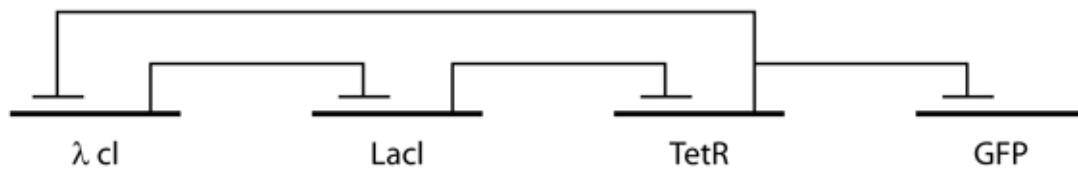


Figure 4. The main structure of the *Repressilator*

The mathematical model Leibler and Elowitz constructed was based on the two differential equations for autorepression. In the case of the *Repressilator*, instead of one gene and its protein, one has 3 genes and proteins—and therefore 6 coupled differential equations of the following form:

$$\frac{dm_i}{dt} = -m_i + \frac{\alpha}{(1 + p^n)} + \alpha_0$$

$$\frac{dp_i}{dt} = -\beta(p_i - m_i)$$

with $\begin{pmatrix} i = lacl, tetR, cl \\ j = cl, lacl, tetR \end{pmatrix}$ .

15

The three proteins *lacl, tetR, cl* are produced by the genes of the *Repressilator*. The set of differential equations is basically of the same form as the one discussed above. It consists of a production and a degradation term. As before $p_i$ denotes the number of proteins and $m_i$ the number of mRNA. In the case of a saturating number of repressors, the number of proteins is given by $\alpha_0$ because of some leakiness at the binding side. In the case of no repressors, the number of proteins is given by $\alpha + \alpha_0 \cdot \beta$, which denotes the ratio of the protein over the mRNA decay rate. The Hill coefficient denoted by $n$ describes the binding strength of the proteins to the transcription site. Thus the differential equations take into account specific biomolecular properties such as leakiness and binding strength.[10] However, those parameters are usually not known and have to be estimated by computer simulations. In those computer simulations a stability diagram is produced marking regions of stable and unstable solutions of the differential equations for different values of $\alpha, \beta$ and $n$. Only when the studied state becomes unstable, sustained oscillations may occur. Since Elowitz and Leibler were interested in regulation by oscillations they focused on the latter case. Only sustained limit-cycle oscillations could provide the rhythm for controlling day and night rhythms in biological organisms.

The computer simulations performed by Leibler and Elowitz gave them more insight into the biochemical conditions of sustained oscillations: "We found that oscillations are favoured by strong promoters coupled to efficient ribosome-binding sites, tight transcriptional repression (low 'leakiness'), cooperative repression characteristics, and comparable protein and mRNA decay rates" (Elowitz and Leibler 2000: 336).

To sum up, the preceding discussion on mathematical modeling in synthetic biology shows how a mathematical model of gene regulation is constructed: by introducing rate equations from chemical kinetics and combining them with a special control mechanism adopted from electrical engineering one can arrive at a general form of coupled differential equations. These differential equations need then to be adapted to the subject matter under investigation by specifying parameters such as binding strength and by exploring different possible dynamics related to the parameters. All these

---

[10] Even if all the active sites of the proteins are occupied by repressors one observes some production of proteins, which is expressed by $\alpha_0$. This is what is meant by leakiness.

modeling activities can be best described as developing and exploring a blueprint for the construction of the subsequent synthetic model.

The *Repressilator*

The synthetic model, the *Repressilator*, was constructed on the basis of the mathematical model and consists of two parts (Figure 5).
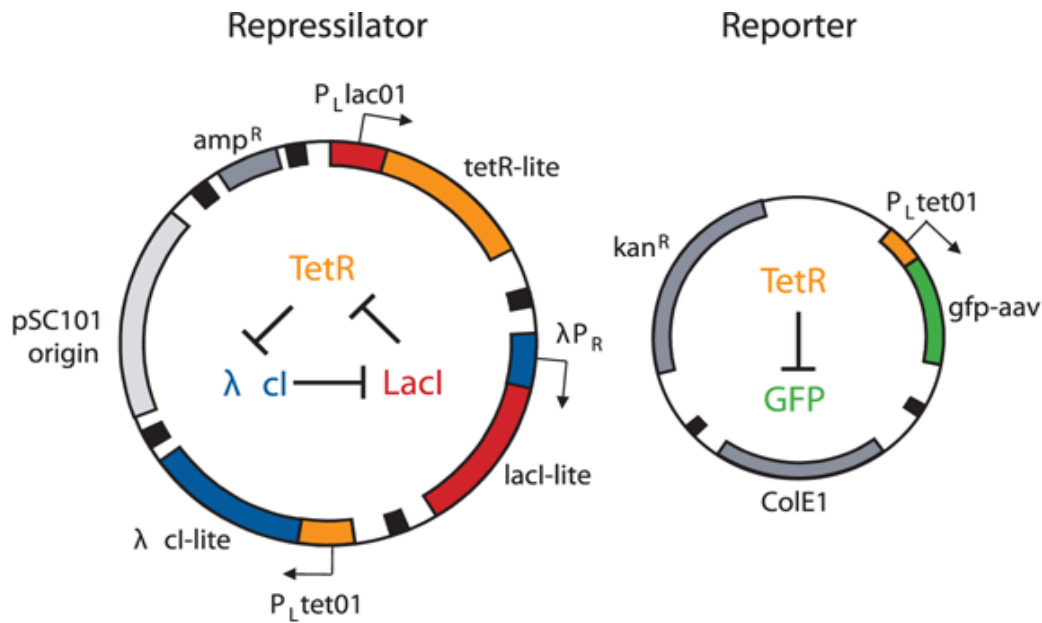


Figure 5. The main components of the *Repressilator* (left hand side) and the *Reporter* (right hand side) (Elowitz and Leibler 2000: 336).

In the diagram the synthetic genetic regulatory network, the *Repressilator,* is shown on the left hand side. The outer part is an illustration of the plasmid constructed by Elowitz and Leibler. The plasmid is an extra-chromosomal DNA molecule integrating the three genes of the *Repressilator*. Plasmids occur naturally in bacteria. In the state of competence, bacteria are able to take up extra chromosomal DNA from the environment. In the case of the *Repressilator,* this property allowed the integration of the specifically designed plasmid into *E.coli* bacteria. The inner part of the illustration represents the feedback loop between the three genes, *TetR*, *LacI*, and λ *cl* , whose dynamics was

studied in advance by the mathematical model. The left-hand side of the diagram shows the *Reporter* consisting of a gene expressing a green fluorescent protein (GFP), which is fused to one of the three genes of the *Repressilator*.

The construction of the *Repressilator* critically depended on the development of new methods and technologies, such as the construction of plasmids, Polymerase Chain Reactions (PCR) and Green Fluorescent Proteins (GFP). GFP became available in the mid-1990s (Chalfie, Tu, Euskirchen, Ward, and Prasher 1994) and very soon also fluorescent proteins with yellow (YFP) and red channels (CFP) were introduced (Elowitz, Surette, Wolf, and Leibler, 1997). By fusing GFPs into a gene regulatory network, implemented within for example E. coli, the expression of genes becomes visible and can be analyzed. Figure (6) shows a picture of "blinking bacteria" from the work of Michael Elowitz (Elowitz, Levine, Siggia, and Swain 2000).
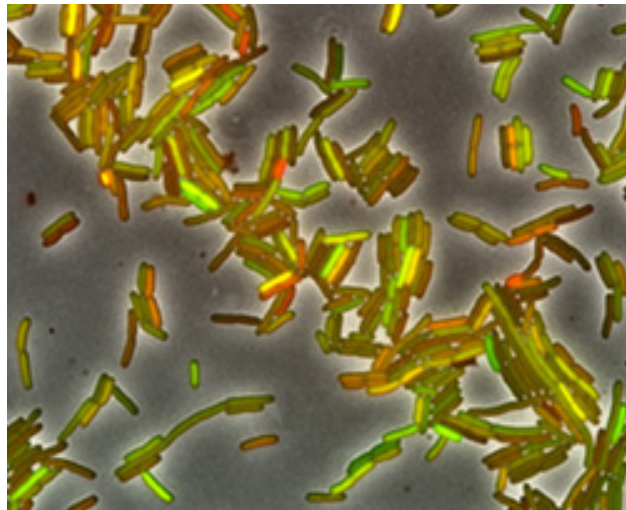


Figure 6. The picture shows E.coli bacteria into which next to GFP's also Yellow and Red Fluorescent Proteins have been introduced (Elowitz, Levine, Siggia, andSwain 2000 1184).

In analyzing the intensity of the light emitted by the GFP, YFP, and CFP of the *E.coli*,

synthetic biologists like Elowitz and his co-workers, try to get insight into the dynamic of such networks and how they give rise to specific biological functions. This kind of analysis comes with several challenges and difficulties. For example, the measurements may indicate that two genes interact, but this does not necessarily mean that one can assign in a straightforward fashion a mechanism underlying that interaction. Moreover, even if the two genes interacted, this does yet not mean that this interaction would play any functional role in the biological system.

The GFP oscillations in the protein level of the *Repressilator* made visible the molecular behavior of transformed cells, i.e. the cells in which the *Repressilator* was implanted. It turned out that the *Repressilator* was indeed able to produce oscillations at the protein level but these oscillations showed irregularities. Interestingly, to find out what was causing such noisy behavior Elowitz and Leibler reverted back to mathematical modeling. In designing the *Repressilator*, Elowitz and Leibler had used a deterministic model. A deterministic model does not take into account stochastic effects such as stochastic fluctuations in gene expression. Performing computer simulations on a stochastic version of the original mathematical model, Elowitz and Leibler were able to reproduce similar variations in the oscillations as observed in the synthetic model. This led researchers to the conclusion that stochastic effects may play a role in gene regulation—which gave rise to a new research program attempting to identify sources of noise in biological systems and the effect of noise on the dynamics of the system (e.g., Swain et al. 2002). This research program makes extensive use of combinational modeling: the role of noise in biological systems was not only studied and explored by making use of mathematical and synthetic modeling but also by comparing the network architectures in model organisms such as *B*. *subtilis* and synthetic systems (e.g., Süel et al. 2007). Model organisms have become an increasingly important part of the modeling practice of synthetic biology laboratories.

## 5. Fictions: Abstract and Concrete

Above we have described the complex interplay of mathematical modeling and synthetic modeling in synthetic biology. Due to the way mathematical models are constructed they

remain abstract and describe only possibilities. However, this is also an advantage of mathematical modeling as the abstract general templates make it possible to study several possible scenarios by adjusting them accordingly. This gives modelers a handle on how things could be and what reasons might underpin why these things might be organized in this or that way. The case of the *Repressilator* showed how synthetic modeling can probe the biological realisticness or implementability of the possible mechanisms depicted by mathematical models.

It is already evident how synthetic modeling has affected synthetic biology: Biology in all its complexity has occupied the central stage. Important engineering notions on which synthetic biology has been grounded, such as noise and modularity, have been reinterpreted and some analogies drawn to engineering have been questioned (see Knuuttila and Loettgers 2013, 2014).

Yet, in order to study the new questions raised by synthetic modeling, researchers typically revert back to mathematical modeling. A good example of this is provided by a recent study by Tatiana T. Marguéz-Lago and Jörg Stelling (2010) who, by employing a series of what they call "minimal models," studied some counter-intuitive behaviors of genetic circuits with negative feedback. As discussed above, the *Repressilator* and related studies made synthetic biologists seriously consider how noise could have a functional role in biological organization (cf. Loettgers 2009). Marguéz-Lago and Stelling have further analyzed the implications of stochastic fluctuations (i.e., noise) by mathematical modeling. They write: "It has often been taken for granted that negative feedback loops in gene regulation work as homeostatic control mechanisms. If one increases the regulation strength a less noisy signal is to be expected. However, recent theoretical studies have reported the exact contrary, counter-intuitive observation, which has left a question mark over the relationship between negative feedback loops and noise" (Marguéz-Lago and Stelling 2011: 1743). Marguéz-Lago and Stelling's article is a telling example of how mathematical models are used to explore different possible explanations for such unexpected behaviors. Starting out from a simple *toy model*, one that cannot represent realistically any biological system, the scientists explore the conditions for different observed behaviors. They create different possible design principles, which could occur but do not necessarily exist in any natural systems. Thus the way

mathematical models are designed and used in synthetic biology serves to highlight their fictional character.

This exploration of possible natural design principles resonates interestingly with the recent philosophical discussion on the fictional nature of modeling (cf. Suárez 2009). For instance, Peter Godfrey-Smith approaches the contemporary model-based theoretical strategy in terms of *imagined non-actual, objects* which are investigated and explored in order to learn something about real-world objects. An important property of these imagined non-actual objects is that they could be concrete if real. Or in the words of Godfrey-Smith: "[…] what I see model-builders are after is trying to describe and understand systems that are only imaginary, but which would be *concrete if real*" (Godfrey-Smith 2009: 101). Synthetic biologists proceed in this way, taking this process even a step further by constructing concrete fictions. The mark of fiction is thus not in its imaginary non-concrete nature but its being a self-contained system that can be manipulated and explored (Knuuttila 2009; Rouse 2009). By engineering gene regulatory networks from biological components synthetic biologists design concrete fictions, which can be tested by and compared with mathematical models—or even transferred into an engineered object fulfilling a specific task.

It is not difficult to uncover the fictional features of a synthetic model such as the *Repressilator* although it is a biological construct functioning in a living cell: Its components (and their number and arrangement) had to be chosen in view of what would be optimal for the behavior under study. The genes used in the *Repressilator* do not occur in such a combination in any known biological system but are chosen and tuned on the basis of the simulations of the underlying mathematical model and other background knowledge[11]—in such a way that the resulting mechanism would allow for sustained oscillations. These technical constraints imply a constraint on what can be explored by such synthetic models: they also study possible design principles in biological systems. In that synthetic models are like mathematical models, they still only provide "how-possibly" explanations. This emphasis is clear  from the writings of synthetic biologists.

---

[11] This draws synthetic modeling close to simulation modeling, which brings to mathematical modeling exploratory and experimental features (e.g., Lenhard 2007).

Leibler and Elowitz did not claim that their synthetic system corresponds to any actual mechanism at work in biological systems. On the contrary, they were very much aware of the limitations of their procedure of drawing analogies to mechanisms, which have been proven to work in engineering but not necessarily in biology. Elowitz and Leibler described their expectations concerning the outcome of the *Repressilator*: "We did not set out to describe precisely the behaviour of the system, as not enough is known about the molecular interactions inside the cell to make such a description realistic. Instead, we hope to identify possible classes of dynamic behaviour and determine which experimental parameters should be adjusted to obtain sustained oscillations" (Elowitz and Leibler 2000: 337). Sprinzak and Elowitz in turn write in the introduction of their review article on synthetic biology: "They [synthetic models] fail to operate as reliably, but they provide a

proof of principle for a synthetic approach to understanding genetic circuits" (Elowitz and Sprinzak 2005: 443). Accordingly, synthetic models could provide a proof of principle for the possibility that such a mechanism as negative feedback could function as control mechanisms in biological systems. This is due to the fact that, despite their fictional character, synthetic models are closer to the actual biological organisms, in so far as they are expected to function under the same material constraints as biological systems. This feature draws synthetic models closer to experimentation and because of this they can be seen as partly bridging the gap between experimentation in model organisms and mathematical modeling. But such a proof is of course far from definite, which is precisely the reason synthetic biologists make use of the combinational approach.

Finally, the fictional nature of synthetic models shows also what goes unrecognized if one takes too literally the idea of mathematical models as *blueprints* for the design and construction of synthetic models. Namely, when talking about mathematical models, synthetic biologists often refer to them as blueprints. Yet the notion of a *blueprint* gives the impression of a ready-made, fixed thing that would function in a more definite manner, like architectural plans for a house. To describe the mathematical model underlying a synthetic model as a blueprint partly misses the explorative role of mathematical models. They provide tools for studying possible

realizations or scenarios, or what synthetic biologists call design principles or motifs, emulating engineering scientists.

## 6. Concluding Remarks

Above we have studied the ways in which synthetic biologists make use of compound analogies by invoking engineering notions such as feedback system, and utilizing computational templates from the study of complex systems. We have argued that because the mechanisms underlying biological functions such as the circadian clock are largely not known, scientists probe them by using control mechanisms, which have been proven to work in other scientific contexts. This makes mathematical modeling, we suggest, inherently fictional (cf. Weisberg 2007) — but it simultaneously enables scientists to make use of cross-disciplinary computational templates and modeling methods. Indeed, the tools and templates that have been developed over the last decades by the study of complex systems provide an important reason why synthetic biologists make use of feedback mechanisms in describing and designing mathematical models of gene regulatory networks. Here also the advancement of computer technologies and the possibility of simulating the non-linear dynamics of feedback systems played a prominent role. Only with the availability of computer technologies and simulations could the space of possible dynamic behaviors of mathematical models be explored, and made use of in the construction of synthetic models.

But of course this analogical procedure of transporting concepts and tools from other fields of study is bound to introduce some uncertainties in the new terrain of application. As we have seen, such engineering-inspired control mechanisms may not resemble those that have evolved in natural processes. They are, indeed, merely possible design principles. This then comes close to the present philosophical discussion on the fictional nature of modeling—moreover by providing a rationale for it, something that the philosophical discussion on fictions largely lacks. This fictional character is also affirmed by synthetic biologists themselves who envision that, as a result of the synthetic approach, the entire field of biology might undergo an important change "from a

discipline that focuses on natural organisms to one that includes potential organisms" (Elowitz and Lim 2010: 889).

In this paper we have concentrated on the basic science-oriented branch of synthetic biology that seeks to understand the general design principles of biological organization on the level of their minimal, or sufficient, components and features. Let us note, however, the double meaning of the quest for design principles in synthetic biology. On the one hand, as we have discussed, synthetic biologists consciously create fictional systems in order to try out various design principles. In electrical engineering, for example, these design principles have well-understood properties and the challenge is to find out whether, and to what extent, they apply in the context of biology. On the other hand, the study of possible design principles aims for engineering *novel* biological parts or systems. Even if such design principles may not have evolved, they could be constructed and used for various purposes, for example for vaccines (e.g., the work of Jay Kiesling).

Last but not least, the testing of whether a design principle can be found to operate in natural systems requires a laborious combinational use of mathematical models, experimentation on model organisms, and synthetic models. This approach has already led to a change in our understanding of how biological systems function and served to underline their differences vis-à-vis engineered artifacts. It will be interesting to see what kind of impact this will have on the mathematical methods and techniques used in modeling biological organization.

## References

Alon, Uri. 2006. *An introduction to systems biology*. London: Chapman & Hall/CRC Mathematical and Computational Biology.
Bangu, Sorin. 2012. *The applicability of mathematics in science: Indispensability and ontology*. Basingstoke: Palgrave Macmillan.
Berg, Jeremy Mark, John L.Tymoczko, and Lubert Stryer. 2002. *Biochemistry*. New York: W. H. Freeman.
Bujara, Matthias, Michael Schümperli, René Pellaux, Matthias Heinemann, and Sven Panke. 2011.Optimization of a blueprint for in vitro glycolysis by metabolic real-time analysis. *Nature Chemical Biology 7*: 271-277.

Çağatay, Tolga, Marc Turcotte, Michael B. Elowitz, Jordi Garcia-Ojalvo, and Gürol M. Suel. 2009. Architecture-dependent noise discriminates functionally analogous differentiation circuits. *Cell* 139(3): 1-11.

Chalfie, Martin, Yuan Tu, Ghia Euskirchen, William W. Ward, and Douglas C. Prasher. 1994. Green fluorescent protein as a marker for gene expression. *Science* 263(5148): 802-805.

Church, George M. 2005. From systems biology to synthetic biology. *Molecular Systems Biology* 1.

Colyvan, Mark. 2001. *The indispensability of mathematics*. New York: Oxford University Press.

Elowitz, Michael B., Michael G. Surette, Pierre-Etienne Wolf, Jeff Stock, and Stanislas Leibler. 1997. Photoactivation turns green fluorescent protein red. *Current Biology* 7(10): 809-812.

Elowitz, Michael B., and Stanislas Leibler. 2000. A synthetic oscillatory network of transcriptional regulators. *Nature* 403(6767): 335-338.

Elowitz, Michael B., Arnold J. Levine, Eric D. Siggia, and Peter S. Swain. 2000. Stochastic gene expression in a single cell. *Science* 297(5584): 1183-1186.

Elowitz Michael B. and Wendell A. Lim. 2010. Build life to understand it. Nature 468(7326): 889-890.

Godfrey-Smith, Peter. 2009. Models and fictions in science. *Philosophical Studies* 143(1): 101-116.

Goodwin, Brian. 1963. *Temporal organization in cells*. London, New York: Academic Press.

Hesse, Mary B. 1966. *Models and analogies in science*. Notre Dame: Notre Dame University Press.

Humphreys, Paul. 2004. *Extending ourselves: Computational science, empiricism, and scientific method*. Oxford: Oxford University Press.

Jacob, François, and Jacques Monod. 1961. Genetic regulatory mechanisms in the synthesis of proteins. *Journal of Molecular Biology* 3(3) 318-356.

Knuuttila, Tarja. 2009. Representation, idealization, and fiction in economics: From the assumptions issue to the epistemology of modelling, in *Fictions in Science: Philosophical essays on modeling and idealization*, ed. Mauricio Suárez, 205-231. New York & London: Routledge.

Knuuttila, Tarja. 2011. Modeling and representing: An artefactual approach. *Studies in History and Philosophy of Science* 42(2): 262–271.

Knuuttila, Tarja and Andrea Loettgers . 2011. The productive tension: Mechanisms vs. templates in modeling the phenomena. In *Representations, Models, and Simulations*, ed. Paul Humphreys and Cyrille Imbert, 3-24. New York: Routledge.

Knuuttila, Tarja and Andrea Loettgers. 2013. Basic science through engineering: Synthetic modeling and the idea of biology-inspired engineering. *Studies in History and Philosophy of Biological and Biomedical Sciences* 44(2): 158–169.

Knuuttila, Tarja and Andrea Loettgers. 2014. Varieties of noise: Analogical reasoning in synthetic biology. *Studies in History and Philosophy of Science Part A* 48:76-88.

Knuuttila, Tarja and Andrea Loettgers. 2016. Modelling as indirect representation? The Lotka-Volterra model revisited. *British Journal for the Philosophy of Science*. doi: 10.1093/bjps/axv055.

Khalil, Ahmad S., and James J. Collins. 2010. Synthetic biology: application come to age. *Nature Reviews Genetics* 11(5): 367-379.

Lenhard, Johannes. 2007. Computer simulation: The cooperation between experimenting and modeling. *Philosophy of Science* 74(2): 176–194.

Loettgers, Andrea. (2009) Synthetic biology and the emergence of a dual meaning of noise. *Biological Theory* 4(4): 340-349.

Marguéz-Lago, Tatiana, and Jörg Stelling. 2010. Counter-intuitive stochastic behavior of simple gene circuits with negative feedback. *Biophysical Journal* 98(9): 1742-1750.

Nandagopal, Nagarajan, and Michael B. Elowitz. 2011. Synthetic biology: Integrated gene circuits. *Science* 333(6047): 1244.

Pincock, Christopher. 2012. *Mathematics and scientific representation*. Oxford and New York: Oxford University Press.

Pittendrigh, Colin S. 1961. On temporal organization in living systems. *The Harvey Lectures* 59: 63-125.

Putman, Hilary. 1975. What is mathematical truth? *Historia Mathematica* 2(4): 529-533.

Rouse, Joseph. 2009. Laboratory fictions. In *Fictions in science: Philosophical essays on modeling and idealization*, ed. Mauricio Suárez, 37-55. New York & London: Routledge.

Sprinzak, David, and Michael B. Elowitz. 2005. Reconstruction of genetic circuits. *Nature* 438(7067): 443-448.

Steiner, Mark. 1998. *The applicability of mathematics as a philosophical problem*. Cambridge, MA: Harvard University Press.

Suárez, Mauricio. 2009. *Fictions in science: Philosophical essays on modeling and idealization*. New York: Routledge.

Süel Gürol M., Rajan P. Kulkarni, Jonathan Dworkin, Jordi Garcia-Ojalvo, and Michael B. Elowitz. 2007 Tunability and noise dependence in differentiation dynamics. *Science* 315(5819): 1716-1719.

Swain, Peter S., Michael Elowitz, and Eric D. Siggia. 2002. Intrinsic and Extrinsic Contributions to Stochasticity in Gene Expression. *Proceedings of the National Academy Sciences* 99(20): 12795-12800.

Thomas, René, and Richard D'Ari. 1990. *Biological feedback*. Boca Raton, Florida, USA: CRC Press.

Weisberg, Michael. 2007. Three kinds of idealization. *The Journal of Philosophy* 104(12): 639-659.

Wigner, Eugene P. 1960. The unreasonable effectiveness of mathematics in the natural sciences. *Communication on Pure and Applied Mathematics* 13(1): 1-14.

Winfree, Arthur. 1967. Biological rhythms and the behavior of populations of coupled oscillators. *Journal of Theoretical Biology* 16(1): 15-42.

Winfree, Arthur. 2001. *The geometry of biological time*. Heidelberg, New York: Springer Verlag.

Zhang, Fuzhong, Sarah Rodriquez, and Jay D. Keasling. 2011. Metabolic engineering of microbial pathways for advanced biofuels production. *Current Opinion in Biotechnology* 22(6): 775-783.